

Listing of the Claims

The listing of claims that follows will replace all prior versions in the application.

1. (Currently Amended) A method of comparing the semantic content of two or more documents, comprising:

accessing two or more documents;

performing a linguistic analysis on each document;
~~outputting a quantified representation of the semantic content of each document; and~~
~~comparing the quantified representations using a defined algorithm.~~

outputting a semantic vector for each document, said semantic vector having

multiple components, wherein each component of said semantic vector has at least:

a word or phrase appearing in the document or a synonym of said word or

phrase;

a weighting factor associated with said word or phrase or synonym; and

a frequency value.
2. (Previously Presented) The method of claim 1, wherein the linguistic analysis comprises sentence analysis.
3. (Previously Presented) The method of claim 2, wherein the sentence analysis comprises a syntactic analysis and a semantic analysis.
4. canceled.
5. canceled

6. (Currently amended) The method of claim 5 1, wherein each component **of the semantic vector** can have multiple dimensions.
7. (Currently amended) The method of claim 6 1, wherein each component of the semantic vector ~~includes one or more text values~~ **further comprises a subordinate concept value.**
8. canceled
9. canceled
10. (Currently amended) ~~The method of claim 8, wherein each component of the semantic vector has two values:~~
~~a word or phrase appearing in the document or a synonym of said word or phrase; and~~
~~a weighting factor associated with said word or phrase or synonym~~
The method of claim 1, wherein some of components of the semantic vector have {main term – subordinate term pairs} as their first value.
11. (Currently amended) The method of claim 4 1, wherein the semantic vector is a multi-dimensional vector defined by the content of a semantic net.

12. (Previously Presented) The method of claim 11, wherein the content of the semantic net is augmented by relative weights, strengths, or frequencies of occurrence of the features within the semantic net.
13. (Previously Presented) The method of claim 1, wherein the output of said defined algorithm is a measure of at least one of semantic distance, semantic similarity, semantic dissimilarity, degree of patentable novelty and degree of anticipation.
14. (Currently Amended) A method of comparing two or more documents, comprising:
linguistically analyzing two or more documents;
generating a semantic vector associated with each document; and
comparing the semantic vectors using a defined metric, **wherein said metric measures the semantic distance between two documents as a function of the relative frequencies of common terms and of common {main term-subordinate term pairs} between the two documents.**
15. (Canceled).
16. (Currently Amended) The method of claim ~~15~~14, wherein a common term between two documents includes two terms that are synonyms.
17. (Previously Presented) The method of claim 14, wherein one or more of said two or more documents are located using an autonomous software or 'bot program.

18. (Previously Presented) The method of claim 17, wherein the 'bot program:
automatically analyzes each document in a defined domain or network by
executing a series of rules and assigning an overall score to the document.
19. (Previously Presented) The method of claim 18, wherein all documents with a score
above a defined threshold are linguistically analyzed.
20. (Previously Presented) The method of claim 14, wherein the semantic vector is a
quantification of the semantic content of each document.
21. (Previously Presented) The method of claim 14, wherein the semantic vector can have
multiple components, and each component can have multiple dimensions.
22. (Previously Presented) The method of claim 14, wherein each component of the
semantic vector has a word or phrase appearing in the document or a synonym of said
word or phrase; and
at least one of a weighting factor associated with said word or phrase or synonym and a
frequency value.
23. (Currently Amended) A system for comparing two or more documents, comprising:
a document inputter, arranged to access two or more documents;
a semantic analyzer, arranged to perform a linguistic analysis on each document;

a semantic quantifier, arranged to output a quantified representation of a semantic content of each document; and

a comparator, arranged to compare the quantified representations using a defined algorithm, **wherein said defined algorithm measures the semantic distance between two documents as a function of the relative frequencies of common terms and of common {main term-subordinate term pairs} between the two documents.**

24. (Currently Amended) A system for comparing two or more documents, comprising:
- a document inputter, arranged to access two or more documents;
 - a semantic analyzer, arranged to perform a linguistic analysis on each document;
 - a semantic vector generator, arranged to output a semantic vector associated with each document; and
 - a comparator, arranged to compare the semantic vectors using a defined metric, **wherein said metric measures the semantic distance between two documents as a function of the relative frequencies of common terms and of common {main term-subordinate term pairs} between the two documents.**

25. (Canceled).

26. (Currently Amended) A computer program product comprising a computer usable medium having computer readable program code means embodied therein, the computer

readable program code means in said computer program product comprising means for causing a computer to:

- access two or more documents;
- perform a linguistic analysis on each document;
- output a quantified representation of a semantic content of each document; and
- compare the quantified representations using a defined algorithm, **wherein said defined metric measures the semantic distance between two documents as a function of the relative frequencies of one of common terms and common {main term-subordinate term pairs} between the two documents.**

27. (Currently Amended) A computer program product comprising a computer usable medium having computer readable program code means embodied therein, the computer readable program code means in said computer program product comprising means for causing a computer to:
- linguistically analyzing two or more documents;
 - generating a semantic vector associated with each document; and
 - comparing the semantic vectors using a defined metric, **wherein said metric measures the semantic distance between two semantic vectors as a function of the relative frequencies of one of common terms and common {main term-subordinate term pairs} between the two documents.**

28. (Previously Presented) The computer program product of claim 27, wherein the computer readable program code means in said computer program product further comprises means for causing a computer to:
identify one or more of said two or more documents using an autonomous software or 'bot program.
29. (Previously Presented) The computer program product of claim 27, wherein said 'bot program automatically analyzes each document in a defined domain or network by executing a series of rules and assigning an overall score to the document.
30. (Previously Presented) The computer program product of claim 27, wherein the semantic vector is a quantification of the semantic content of each document.
31. (Previously Presented) The computer program product of claim 27, wherein the output of said defined metric is a measure of at least one of semantic distance, semantic similarity, semantic dissimilarity, degree of patentable novelty and degree of anticipation.
32. (Canceled).
33. (New) A system for comparing two or more documents, comprising:
a document inputter, arranged to access two or more documents;
a semantic analyzer, arranged to perform a linguistic analysis on each document;

a semantic vector generator, arranged to output a semantic vector associated with each document; and

a comparator, arranged to compare the semantic vectors using a defined metric, wherein said defined metric is one of:

$$[\text{Sqrt}(f_1^2 + f_2^2 + f_3^2 + f_4^2 + \dots + f_{(N-1)}^2 + f_N^2)/n] * 100,$$

wherein f is a difference in frequency of a common term between two documents and n is the number of terms those documents have in common; or

$$\text{Sqrt}(\text{sum}((w-\text{Delta})^2 * w-\text{Avg})) / (\text{Log}(n)^3 * 1000),$$

wherein w-Delta is the difference in weight between two common terms, w-Avg is the average weight between two common terms, and n is the number of common terms, between two documents.

34. (New) A method of comparing two or more documents, comprising:

linguistically analyzing two or more documents;

generating a semantic vector associated with each document; and

comparing the semantic vectors using a defined metric, wherein said defined metric is one of:

$$\text{Sqrt}(f_1^2 + f_2^2 + f_3^2 + f_4^2 + \dots + f_{(N-1)}^2 + f_N^2)/n] * 100$$

wherein f is a difference in frequency of a common term between two documents and n is the number of terms those documents have in common; or

$$\text{Sqrt}(\text{sum}((w\text{-Delta})^2 * w\text{-Avg})) / (\text{Log}(n)^3 * 1000),$$

wherein w-Delta is the difference in weight between two common terms, w-Avg is the average weight between two common terms, and n is the number of common terms, between two documents.

35. (New) A computer program product comprising a computer usable medium having computer readable program code means embodied therein, the computer readable program code means in said computer program product comprising means for causing a computer to:
- access two or more documents;
 - perform a linguistic analysis on each document;
 - output a quantified representation of a semantic content of each document; and
 - compare the quantified representations using a defined algorithm, wherein said defined algorithm is one of:

$$[\text{Sqrt}(f_1^2 + f_2^2 + f_3^2 + f_4^2 + \dots + f_{(N-1)}^2 + f_N^2)/n] * 100,$$

wherein f is a difference in frequency of a common term between two documents and n is the number of terms those documents have in common; or

$$\text{Sqrt}(\text{sum}((w\text{-Delta})^2 * w\text{-Avg})) / (\text{Log}(n)^3 * 1000),$$

wherein $w\text{-Delta}$ is the difference in weight between two common terms, $w\text{-Avg}$ is the average weight between two common terms, and n is the number of common terms, between two documents.